

## Huygens ING herschrijft de grondwet van XML

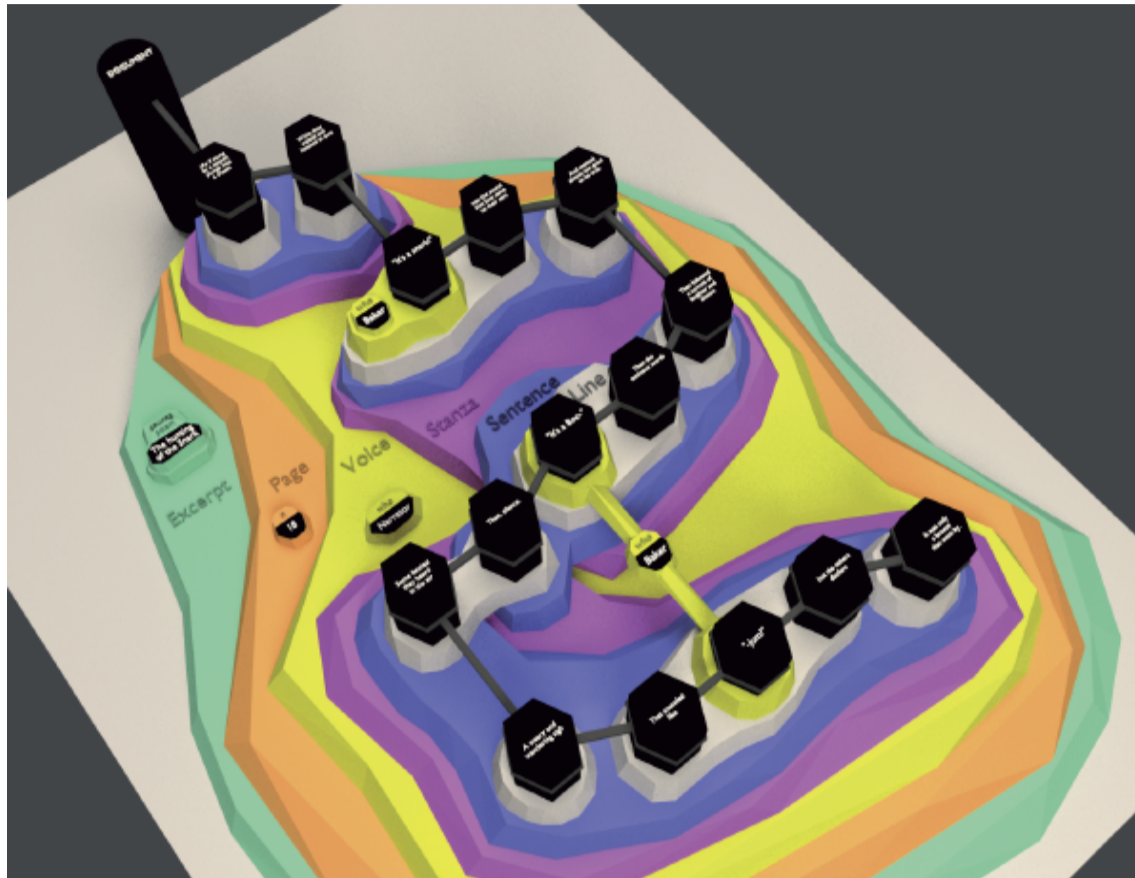
Thijs van der Veen

Ronald Haentjens Dekker, lead engineer van de R&D-afdeling van het Huygens ING, presenteerde begin augustus een paper op de Balisage: The Markup Conference 2017 in Washington D.C. die insloeg als een bom. In zijn revolutionaire paper stellen hij en co-auteur David Birbaum dat, om tekstdocumenten te modelleren en te verwerken, je tekst veel beter kan benaderen als een graph (Text as graph = TAG), dan als een boomstructuur, zoals nu in XML het geval is. "Het XML-boomparadigma heeft verschillende bekende beperkingen voor documentmodellering en -verwerking die op deze manier bijna allemaal worden opgelost," zegt Haentjens Dekker. Michael Sperberg-McQueen, Founding Father van XML, noemde het een briljant paper.

Gertjan Filarski, hoofd van Haentjens Dekkers afdeling, zegt: "Het heeft de potentie om de eerste fundamentele wijziging in SGML/XML te zijn in meer dan 25 jaar." Je kan het qua impact vergelijken met een grondwetswijziging: het komt zelden voor en het gaat niet over één nacht ijs. Dit nieuwe model kan in theorie een enorme impact hebben op ons dagelijks leven want XML ligt namelijk aan de basis van websites en tekstverwerkingsprogramma's zoals Word.

Op basis van het TAG-model wil het Huygens ING een nieuwe *markup language* ontwikkelen. In tegenstelling tot XML kan deze taal meerdere informatielagen tegelijkertijd vastleggen. Dit biedt wetenschappers de mogelijkheid om vele nieuwe onderzoeksvragen te formuleren. Binnen vijf jaar kan deze *markup language* in de digitale research tools van het Huygens ING geïmplementeerd zijn. Dit betekent trouwens niet automatisch dat onderzoekers op korte termijn zich deze nieuwe taal ook eigen moeten maken. Balisage is een jaarlijkse conferentie gewijd aan de theorie en praktijk van *descriptive markup* en

## GEHOORD & BIJGEWOOND



Deze illustratie is een schematische weergave van hoe TAG de verschillende informatielagen op dezelfde verzameling tekst-nodes vastlegt. De presentatie hierover tijdens Balisage: The Markup Conference 2017 sloeg in als een bom. credits Gijsjan Brouwer, Huygens ING

aanverwante technologieën voor het structureren en beheren van informatie.

[huygens.knaw.nl](http://huygens.knaw.nl)

### ICPSR Summer Program: leren, zien en genieten

Kasper Otten

Afgelopen zomer heb ik, met behulp van een subsidie van DANS, deelgenomen aan het ICPSR Summer Program 2017 in Ann Arbor, Michigan. ICPSR staat voor Inter-university Consortium for Political and Social Research, hun jaarlijkse zomerschool is wereldwijd toonaangevend in cursussen op het gebied van kwantitatieve methoden van sociaalwetenschappelijk onderzoek. Ik heb daar twee cursussen gevolgd en aanvullende collegereeksen bezocht. De eerste cursus was Game Theory II: Advanced

Topics, waarin theoretische modellen voor strategische interactie tussen personen worden behandeld. De tweede cursus was Causal Inference for the Social Sciences, waarin statistische methodes worden behandeld om causale verbanden vast te stellen. De collegereeksen behandelden software om de statische methodes concreet toe te passen. Theorie, statistiek en software in één zomerprogramma dus. Dat klinkt als veel, en dat was het ook wel. Voor elk onderdeel was huiswerk te maken, en ook al was het niet verplicht, ik wilde het toch graag allemaal doen om zoveel mogelijk uit het programma te halen. Dat wil niet zeggen dat er helemaal geen vrije tijd was. Zo kon ik ondertussen nog deelnemen aan sociale activiteiten georganiseerd door het ICPSR, zwemmen in een

nabijgelegen rivier, bezienswaardigheden in Ann Arbor bezoeken en Detroit verkennen. Bovendien heb ik na afloop samen met een studiegenoot een auto gehuurd voor een week, om van Ann Arbor naar New York te rijden via Morgantown, Washington, Philadelphia, Atlantic City en Princeton. Het zal je niet verbazen dat die reis goed beviel na al het huiswerk in de weken ervoor. Kortom, ik heb veel geleerd, gezien, en genoten.

[icpsr.umich.edu](http://icpsr.umich.edu)

### DH2017: Acces/Access

Steven Claeysens

De jaarlijkse mondiale Digital Humanities-conferentie vond deze zomer plaats in Montréal en de Canadezen pakten uit met een primeur: de eerste officieel tweetalige editie. Niet alle deelnemers waren over-

tuigd van de logica, maar de redenering luidde dat meertaligheid de toegankelijkheid van het congres ten goede komt en daarom was behalve het Engels ook het Frans een voertaal. Het Access/Access-thema accentueerde de nieuwe Babelse opzet. Dat thema stond ook centraal in twee van de drie keynotes. In de openingslezing brak Marin Dacos, grote man achter belangrijke Open Access-initiatieven als *revues.org* (tijdschriften), *openedition.org* (monografieën) en *hypotheses.org* (weblogs), een Franse lans voor het belang van Open Science en Open Access (OA). Hij vroeg in het bijzonder aandacht voor *the unexpected reader*, de lezer die niet tot de primaire doelgroep behoort maar wel wetenschappelijke publicaties leest doordat ze vrij toegankelijk online staan. Dacos deed een poging zijn pleidooi kracht bij te zetten door middel van logfile-analyses van de verschillende projecten in zijn portefeuille. De slotlezing van Elizabeth Guffey stond volledig in het teken van de ontoegankelijkheid van grote delen van de digitale ruimte voor mensen met visuele, auditieve of cognitieve beperkingen. Gelukkig kon ze ook wijzen op waardevolle initiatieven, zoals Teach Access en de Accessibility Toolkit van de Kahn Academy. De Zampolli-lezing, behorende bij de gelijknamige prijs, behandelde 30 jaar Text Encoding Initiative (TEI) en werd gehouden door drie spilfiguren uit de TEI-gemeenschap: Nancy Ide, Michael Sperberg-McQueen en Lou Burnard. Naast de plenaire sessies waren er de gangbare workshops, lezingen, parallelsessies, posters en demonstraties – voor elk wat wils en voor niemand allemaal bij te wonen. Anderen zagen wellicht andere zaken, mij sprongen twee trends in het oog: de onstuitbare opmars van de computationele beeldherkenning en de snelle groei van IIF, vaak gecombineerd met Linked Data. Volgend jaar is Mexico City het toneel, het jaar nadien Utrecht.

[dh2017.adho.org](http://dh2017.adho.org)

## RDA-bouwstenen online beschikbaar

Vervolg van pagina 1

“Door gebruik te maken van de kennis en ervaring van haar gemeenschap en betrokken gebruikers, probeert de RDA het delen van data zo concreet mogelijk te maken en te bevorderen.”

### Internationale impact

De eindresultaten staan op de RDA-site: er zijn nu zo'n kleine twintig

zogenoemde RDA Recommendations and Outputs beschikbaar voor hergebruik. Eén hiervan is de CoreTrustSeal-standaard (zie kader). Dillo: “Het is de betrokken werkgroep gelukt om procedures voor het certificeren van repositories van twee keurmerken (WDS en DSA) samen te brengen tot één nieuw mondiaal geldend datakeurmerk: het CoreTrustSeal (CTS). Dit nieuwe keurmerk kan bijvoorbeeld opgeno-

men worden binnen de rules of engagement van de European Open Science Cloud (EOSC). Dit Europese initiatief moet zorgen voor een infrastructuur die het makkelijker maakt voor onderzoekers om data te delen en te combineren, ook over disciplinaire en geografische grenzen heen. De impact van het CTS voor de internationale onderzoeksgemeenschap is groot: met eenduidige, samenhangende normen en standaarden voor repositories groeit ook het vertrouwen van onderzoe-

kers en onderzoeksfinanciers in de toekomstbestendigheid van deze cruciale bewaarplaatsen.”

### Nationaal knooppunt

Een ander voorbeeld van een RDA Recommendation is 23 Things: een gratis online training voor data-managers om (het potentieel van) onderzoeksdata beter te begrijpen. De basisgedachte van 23 Things is dat iedereen op elk moment 23 dingen kan doen op het gebied van onderzoeksdata. De informatie is

met hulp van de RDA inmiddels in 11 talen online beschikbaar.

Dillo sluit af: “Volgend jaar start het Horizon2020 project RDA Europe 4.0. Met de financiering van dit project zet de Europese Commissie haar steun aan de RDA voort. Binnen dit project zal DANS als nationaal knooppunt een flinke boost kunnen geven aan *RDA Netherlands*. Voor wie niet kan wachten: een persoonlijk lidmaatschap is gratis en iedereen is welkom.”

[rd-alliance.org](http://rd-alliance.org)

### OVERNEMEN ARTIKELEN

Wilt u een artikel uit dit blad overnemen?

Dat mag altijd, maar vermeld wel de bron (E-data & Research) en de naam van de auteur van het artikel. Neem ook contact op met de hoofdredacteur (zie colofon) om door te geven waar artikelen geplaatst worden.